PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-232693
(43)Date of publication of application : 02.09.1998

(51)Int.Cl. G10L 3/00
G10L 3/00

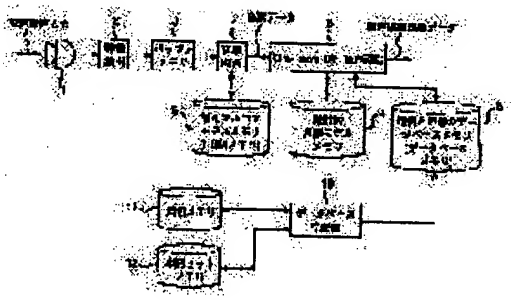
(21)Application number : 09-161243
(22)Date of filing : 18.06.1997

(71)Applicant : ATR ONSEI HONYAKU TSUSHIN KENKYUSHO:KK
(72)Inventor : KAWAI ATSUSHI
WAKITA YUMI

(30)Priority
Priority number : 08341084 Priority date : 20.12.1996 Priority country : JP

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a voice recognition device capable of removing an unsuitable erroneous recognition result, outputting a locally and perspectively suitable sentence and obtaining a voice recognition rate higher than a usual example.
SOLUTION: This device is provided with a voice recognition part 6 recognizing voice referring to a statistical language model based on a voice signal of a voice of an utterance voice sentence consisting of an inputted word line. The voice recognition part 6 calculates a function value of an unsuitable sentence decisional function for a voice recognition candidate by using a prescribed unsuitable sentence decisional function showing an unsuitable extent for the voice recognition candidate, and when the calculated function value exceeds a threshold value, the voice recognition part 6 removes its voice recognition candidate to voice recognize. The function value is made the value that e.g. the sum of meaning distances answering to examples used in voice recognition processing, and the calculated sum is multiplied by the number of form elements incorporated in the voice recognition candidate becoming the subject of the voice recognition processing, and the value is divided by the number of examples used for the voice recognition processing.



LEGAL STATUS

[Date of request for examination]	18.06.1997
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	2965529
[Date of registration]	13.08.1999
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision of rejection]	

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報**（A）　(11)特許出願公開番号

特開平10－232693

(43)公開日　平成10年(1998)９月２日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５３１	Ｇ１０Ｌ　3/00	５３１Ｐ
	５６１		５６１Ｇ

審査請求　有　　請求項の数６　ＯＬ　（全 10 頁）

(21)出願番号　特願平9－161243

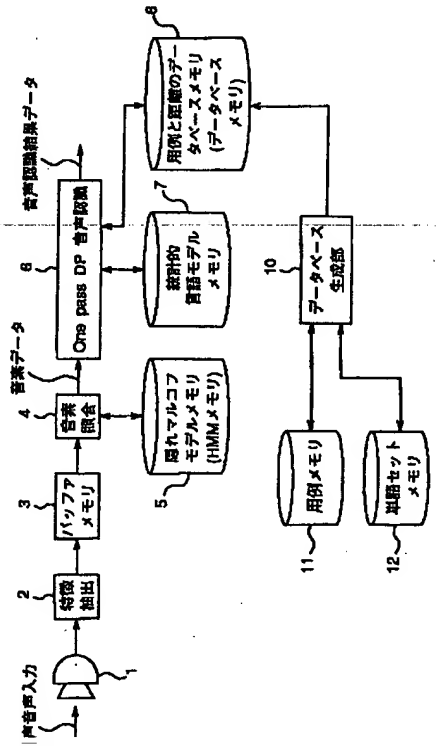
(22)出願日　平成９年(1997)６月18日

(31)優先権主張番号　特願平8－341084
(32)優先日　平８(1996)12月20日
(33)優先権主張国　日本（ＪＰ）
特許法第30条第１項適用申請有り　1996年６月24日　社団法人人工知能学会発行の「1996年度人工知能学会全国大会（第10回）論文集」に発表

(71)出願人　593118597
株式会社エイ・ティ・アール音声翻訳通信研究所
京都府相楽郡精華町大字乾谷小字三平谷５番地
(72)発明者　河井　淳
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内
(72)発明者　脇田　由実
京都府相楽郡精華町大字乾谷小字三平谷５番地　株式会社エイ・ティ・アール音声翻訳通信研究所内
(74)代理人　弁理士　青山　葆　（外２名）

(54)【発明の名称】　音声認識装置

(57)【要約】
【課題】　不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供する。
【解決手段】　入力される単語列からなる発声音声文の音声の音声信号に基づいて、統計的言語モデルを参照して音声認識する音声認識部６を備え、音声認識部６は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値がしきい値を超えるときに、当該音声認識候補を除去して音声認識する。関数値は、例えば、音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。



1

【特許請求の範囲】

【請求項１】　入力される単語列からなる発声音声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、

上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識処理を実行することを特徴とする音声認識装置。

【請求項２】　上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする請求項１記載の音声認識装置。

【請求項３】　上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする請求項１記載の音声認識装置。

【請求項４】　上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値であることを特徴とする請求項１記載の音声認識装置。

【請求項５】　上記しきい値は、一定値であることを特徴とする請求項１乃至４のうちの１つに記載の音声認識装置。

【請求項６】　上記しきい値は、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させることを特徴とする請求項１乃至４のうちの１つに記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、発声音声の音声信号に基づいて、統計的言語モデルを参照して音声認識する音声認識装置に関する。

【０００２】

【従来の技術】連続音声認識装置において、Ｎ－ｇｒａｍと呼ばれる統計的手法に基づいた統計的言語モデルが広く使用されている（例えば、従来技術文献１「L. R. Bahl et al., "A Maximum Likelihood Approach to Continuous Speech Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 179-190, 1983年」参照。）。Ｎ－ｇｒａｍを用いた連続音声認識装置では、予め、大規模な学習データを用いて、直前のＮ－１個の単語から次の単語に遷移する遷移確率を学習しておき、音声認識時に、学習した遷移確率を用いて次に接続する単語を予測することにより、音声認識率の向上を計っている。一般に、Ｎが大きくなるほど次単語の予測精度は向上するが、単語連鎖の種類数が多くなるため、信頼できる遷移確率を得るためには、大量の学習データが必要となる。そこで現状では、Ｎを２（ｂｉ－ｇｒａｍ）又は３（ｔｒｉ－ｇｒａｍ）程度に設定して使用している例が多い。しかしながら、単語のｂｉ－ｇｒａｍや単語のｔｒｉ－ｇｒａｍを用いた連続音声認識結果を分析してみると、２又は３単語以内の局所的な単語連鎖に自然性はあったとしても、文全体を眺めると、不自然な誤認識文を結果として出力している例が多々見受けられ、より大局的な言語制約が必要であると考える。

【０００３】文脈自由文法などの文法や単語間の依存関係を用いて、より大局的な制約を可能とする言語モデルが提案されている。しかしながら、自然発話文の構造や依存関係の多様性を考えると、規則や依存関係の構築は容易ではないし、処理量も膨大になる。一方、用例主導型のアプローチで文の構文の曖昧性を解消する方法（以下、従来例という。）が従来技術文献２「隅田英一郎ほか，“英語前置語句係り先の用例主導あいまい性解消”，電子情報通信学会論文誌（Ｄ－ＩＩ），Ｊ７７－Ｄ－ＩＩ，Ｎｏ３，ｐｐ．５５７－５６５，１９９４年３月」において提案されている。この従来例の方法は、コーパスから用例を抽出し、入力文の表現と用例との意味的距離をシソーラスに従って計算し、最終的な意味的距離が最も小さくなる構文を選択する方法であり、対訳決定処理などでもその効果が確認されている（従来技術文献３「古瀬蔵ほか，“経験的知識を活用する変換主導型機械翻訳”，情報処理学会論文誌，Ｖｏｌ．３５，Ｎｏ３，ｐｐ．４１４－４２３，１９９４年３月」参照。）。

【０００４】

【発明が解決しようとする課題】しかしながら、従来例の方法を用いる音声認識装置において、例えば、学習した用例に対して不自然な構文を入力すると、どの用例との意味的距離も大きくなってしまい、音声認識率が比較的低いという問題点があった。

【０００５】本発明の目的は以上の問題点を解決し、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することにある。

【０００６】

載の音声認識装置は、入力される単語列からなる発声音声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識処理を実行することを特徴とする。

【０００７】また、請求項２記載の音声認識装置は、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする。さらに、請求項３記載の音声認識装置は、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値であることを特徴とする。またさらに、請求項４記載の音声認識装置は、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値であることを特徴とする。

【０００８】また、請求項５記載の音声認識装置は、請求項１乃至４のうちの１つに記載の音声認識装置において、上記しきい値は、好ましくは、一定値である。さらに、請求項６記載の音声認識装置は、請求項１乃至４のうちの１つに記載の音声認識装置において、上記しきい値は、好ましくは、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させる。

【０００９】

【発明の実施の形態】以下、図面を参照して本発明に係る実施形態について説明する。

【００１０】図１は、本発明に係る一実施形態の音声認識装置の構成を示すブロック図である。この音声認識装置は、マイクロホン１と、特徴抽出部２と、バッファメモリ３と、入力される発声音声データに基づいて隠れマルコフモデルメモリ（以下、HMMメモリという。）５

Mという。）を参照して音素照合処理を実行して音素データを出力する音素照合部４と、音素照合部４からの音素データに基づいてOne pass DP（Viterbi search）アルゴリズムを用いて統計的言語モデルメモリ７内の統計的言語モデル及び用例と距離のデータベースメモリ（データベースメモリという。）８内の用例と距離のデータベース（以下、データベースという。）を参照して音声認識を実行するOne pass DP音声認識部（以下、音声認識部という。）６とを備え、上記音声認識部６は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数（詳細後述する数１）を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値Ｆｔｈを超えるときに、当該音声認識候補を除去して音声認識することを特徴とする。ここで、上記不適格文判定関数の関数値は、好ましくは、上記音声認識候補の構文を決定するために用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識候補の構文を決定するために用いた用例の数で除算した値である。また、上記しきい値Ｆｔｈは、好ましくは、一定値、又は、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させる。なお、形態素とは、語幹、接頭辞、接尾辞など意味を有する文字系列の最小単位で単語と実質的に同一であるかやや小さい単位である。

【００１１】まず、音声認識部６における不適格文検出手法について説明する。Ｎ－ｇｒａｍを用いた音声認識処理における誤認識には次の特徴がある。

（ａ）Ｎ個以上の単語の連鎖で判断すると、文法的及び意味的に不適当な単語の組み合わせが存在する。例えば、誤認識例：「電話番号が２１０７号室ですか」。

（ｂ）文の構造が大きな単位でまとまらない。すなわち、文法的に規則を適用することができず、局所的にしか判断できない。例えば、誤認識例：「三名様までのえーまでシングルの一泊の……」。

【００１２】上記の特徴を持つ誤認識を解決するためには、Ｎ－ｇｒａｍよりも、より大局的な立場で、単語間の整合性や構文の適格性を判断する必要がある。一方、用例主導型の音声翻訳手法（従来技術文献２及び従来技術文献４「O. Furuse et al., "Incremental Translation Utilizing Constituent Boundary Patterns", Proceedings of Coling'96, 1996年」参照。）では、用例に基づく翻訳知識を用いて左から右に方向で（left-to-rightに）構文を決定していく方法をとっている。この処理課程で、構文の曖昧性を解消するために、入力文と用例との意味的距離をシソーラス（類語辞書）を用いて計算し、距離の小さい用例に相当する構文を選択する方法をとっている。本発明者は、次の理由により、上記構文決

識を除去するのに整合性が良いと考えられる。

（ａ）上記構文決定手法は用例主導型手法であるので、会話文に見られるような従来の文法で処理が困難な構文が容易に処理可能である。

（ｂ）上記構文決定手法では、構文に基づいて意味的距離を求めているので、隣接しない単語間の整合性を判断できる能力がある。

（ｃ）音声認識、上記構文決定手法、ともに左から右に方向で（left-to-rightに）処理を行なっているので、ある時点までの中間結果を、逐次的に判定できる可能性がある。

【００１３】そこで、大局的にみた意味的距離の整合性と解析された構文の適格性で、不適格文を検出する。具体的には次のように判断する。まず、部分文における意味的距離の不整合は、上記の構文決定手法に用いた意味的距離値で判断する。ある部分文の意味的距離の総和が一定値以上になると、その文を誤認識と判断する。次に構文の適格性については次のように考える。一定以上の形態素からなる自然な文であればまとまった構文を持ち、構文の構造はある程度複雑な構造であろうと仮定する。ある部分文に含まれる形態素の数mの、構文決定のために使用された文脈自由文法の規則又は用例の規則数（又は用例数）Ｒに対する割合（＝m／Ｒ）を考える。まとまった構文を持たない部分文は構文構造が階層にならず、よって形態素の数mに対して、使用された構文規則数Ｒは少なく、m／Ｒ値は大きくなる。逆に、構文が複雑になり階層的になるほど、m／Ｒ値は小さくなる。そこで、次式の不適格文判定関数Ｆerror（m）を定義する。

【００１４】

【数１】

$$F_{error}(m) = (m/R) \sum_{i=1}^{R} d(r_i)$$

【００１５】ここで、ｄ（r_i）は複数の用例又は規則r_iに対応する意味的距離又は類似度距離であり、mは音声認識処理の対象となる音声認識候補の部分文に含まれる形態素の数であり、Ｒは音声認識処理を実行するとき音声認識候補の部分文の構文を決定するために用いた用例又は規則の数である。ここで、意味的距離又は類似度距離とは、例えば従来技術文献２のｐ．５５９の（１）式で定義され、シソーラスを用いて計算する、入力発声音声文の音声認識候補と用例との間の意味的距離であって、本実施形態においては、音声認識候補の部分文に該当するデータベース内の用例に対する距離を検索

して決定する。ここで、シソーラスとは、概念間の上位下位関係を木構造で表現し、葉に相当する最下位の概念に当該概念をもつ単語を割り当てた辞書を指す。単語間の意味的距離はシソーラス上の概念間の意味的距離によって定義され、概念間の距離はシソーラスにおける最小の共通上位概念の位置に従って０から１までの値に設定される。値０は２つの概念が同じであることを意味し、値１は無関係であることを意味する。また、上記判定関数Ｆerror（m）は形態素数mの関数であり、文章の始めからm番目の形態素までの音声認識候補の部分文を対象に計算される。この判定関数値Ｆerror（m）が所定のしきい値Ｆｔｈを越えた場合、音声認識部６は、その音声認識候補の部分文を誤認識結果と判断して、音声認識候補から除去する。なお、上記数１は、好ましくは、m≧５のときに適用することができる。なお、上記数１における規則数Ｒが０であるときは、当該関数値を１とし、誤認識結果と判断して、音声認識候補から除去する。

【００１６】図１の好ましい実施形態においては、データベース生成部１０は、用例メモリ１１内の用例と、単語セットメモリ１２内の単語セットとに基づいて、所定の類似度規則を用いて、データベースを生成して、データベースメモリ８に記憶する。文脈自由文法規則の用例の一例を表１及び表２に示す。また、類似度規則の一例を表３に示す。

【００１７】

【表１】用例１

XのY
僕の子供
あなたの会社
.....

【００１８】

【表２】用例２

XがY
僕が先生
.....

【００１９】

【表３】

類似度規則

（Ｉ）単語セットの組で生成される文が用例と同じとき、距離＝０とする。

（II）単語セットの組で生成される文が用例と同じ機能単語（例えば、

を有するとき、距離＝10^{-5}とする。
（iii）単語セットの組で生成される文が用例に無い単語同士のとき、
距離＝０．５とする。

【００２０】日本語処理の音声認識装置における、単語セットＳ１，Ｓ２，Ｓ３，Ｓ４の一例、並びに、単語セット間の所定の機能単語を用いたときの距離を図２に示す。図２において、例えば、「あなた」（単語セットＳ１）が「先生」（単語セットＳ２）のとき、距離が10^{-5}になり、「あなた」（単語セットＳ１）の「子供」（単語セットＳ３）のとき、距離が10^{-5}になり、「あなた」（単語セットＳ１）の「会社」（単語セットＳ４）のとき、距離が10^{-5}になる。また、「先生」（単語セットＳ２）の「会社」（単語セットＳ４）のとき、距離が０．５になる。

【００２１】データベース生成部１０は、表１及び表２の用例と、表３の類似度規則を用いたときのデータベース生成処理を以下のように行う。各単語セットの組で部分文を生成して、部分文が「あなたの会社」であるときは、距離は０となり、部分文が「私の学校」であるときは、距離は10^{-5}となり、部分文が「子供が先生」であるときは、距離は０．５となる。このように生成した、部分文の用例と距離とのデータベースは、データベースメモリ８に記憶される。

【００２２】さらに、統計的言語モデルは、発声音声文のテキストデータに基づいて、公知の方法により、例えば、単語のｂｉ－ｇｒａｍの統計的言語モデルを生成して統計的言語モデルメモリ７に記憶する。

【００２３】次いで、本実施形態の統計的言語モデルを用いた音声認識装置の構成及び動作について説明する。

【００２４】図１において、話者の発声音声はマイクロホン１に入力されて音声信号に変換された後、特徴抽出部２に入力される。特徴抽出部２は、入力された音声信号をＡ／Ｄ変換した後、例えばＬＰＣ分析を実行し、対数パワー、１６次ケプストラム係数、Δ対数パワー及び１６次Δケプストラム係数を含む３４次元の特徴パラメータを抽出する。抽出された特徴パラメータの時系列はバッファメモリ３を介して音素照合部４に入力される。音素照合部４に接続されるＨＭＭメモリ５内のＨＭＭは、複数の状態と、各状態間の遷移を示す弧から構成され、各弧には状態間の遷移確率と入力コードに対する出力確率を有している。音素照合部４は、入力されたデータに基づいて音素照合処理を実行して音素データを、音声認識部６に出力する。

【００２５】統計的言語モデルを予め記憶する統計的言語モデルメモリ７は音声認識部６に接続される。音声認識部６は、統計的言語モデルメモリ７内の統計的言語モデル及びデータベースメモリ８内のデータベースを参照して、所定のＯｎｅ　ｐａｓｓ　ＤＰアルゴリズムを用後戻りなしに処理してより高い生起確率の単語を音声認識候補として認識し、当該音声認識候補に対して上記数１を用いて判定関数値Ｆｅｒｒｏｒ（ｍ）を計算する。ここで、数１におけるｄ（ｒｉ）は音声認識候補に該当する用例をデータベースより検索して、検索された用例に該当する距離を意味的距離とする。そして、計算された判定関数値Ｆｅｒｒｏｒ（ｍ）が所定のしきい値ＦｔｈをＦｔｈを越えた場合、音声認識部６は、その音声認識候補の部分文を誤認識結果と判断して、音声認識候補から除去する。そして、残った音声認識候補を音声認識結果（文字列データ）と決定して出力する。

【００２６】図３は、以上のように構成された日本語処理の音声認識装置の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。図３（ａ）に示すように、「私のエットー学校がね」という入力文の音声が入力されたとき、認識結果文として、図３（ｂ）に示すように、「私の江藤学校がね」が得られたとき、すなわち、「エットー」という間投詞が「江藤」という名詞に誤って認識された場合である。認識結果文における単語間のスコアを図３（ｂ）に示している。さらに、認識結果文に基づいて構文解析したときに、図３（ｃ）に示すように、より小さいスコアに基づいて構文解析結果の構文木が得られ、このときのスコアが得られている。図３（ｃ）における場合を、上記数１に当てはめると、不適格文判定関数の関数値Ｆｅｒｒｏｒ（ｍ）は次式のようになる。

【００２７】
【数２】

$$\begin{aligned}&Ferror(m)\\&=(6/3)(0.5+0.5+10^{-5})\\&=2\times(1.00001)\\&=2.00002\end{aligned}$$

【００２８】当該例において、不適格文を判定するときのしきい値Ｆｔｈは、好ましくは、０．６乃至０．７であり、上記数２で計算された関数値＝２．００００２はしきい値ＦｔｈをＦｔｈを超えているので、それに対応する音声認識候補は音声認識候補から除去される。上記しきい値Ｆｔｈは、一定値であってもよいし、音声認識処理の対象となる部分文に含まれる形態素数ｍに依存して変化してもよい。

【００２９】以上のように構成された音声認識装置において、特徴抽出部２と、音素照合部４と、音声認識部６と、データベース生成部１０とは、例えば、ディジタル計算機などのコンピュータで構成され、バッファメモリ

と、データベースメモリ８とは、例えば、ハードディスクメモリなどの記憶装置で構成される。

【００３０】次いで、英語処理の音声認識装置の一例について説明する。英語処理のときの文脈自由文法規則の用例の一例を表４及び表５に示す。また、類似度規則は例えば、表３のものをそのまま使用する。

【００３１】

【表４】用例１１

X at Y
start at 7：30
leave at 6 p. m.
.....

【００３２】

【表５】用例１２

Z・X
the train starts
.....

【００３３】英語処理の音声認識装置における、単語セットＳ１１（Ｘ），Ｓ１２，Ｓ１３（Ｚ），Ｓ１４（Ｙ）の一例、並びに、単語セット間の所定の機能単語を用いたときの距離を図５に示す。図５において、例えば、「ｔｒａｉｎ　ｌｅａｖｅｓ」のとき距離が10^{-5}になり、「ｌｅａｖｅ　ｔｒａｉｎ」のとき距離が０．５になる。また、「ｌｅａｖｅ　Ｋｙｏｔｏ」のとき距離が10^{-5}になり、「ｌｅａｖｅ　ａｔ　６　ｐ．ｍ．」のとき距離が10^{-5}になる。データベース生成部１０は、表４及び表５の用例と、表３の類似度規則を用いたときのデータベース生成処理を以下のように行う。各単語セットの組で部分文を生成して、部分文が「ｔｈｅ　ｔｒａｉｎ　ｓｔａｒｔｓ」であるときは、距離は０となり、部分文が「ｔｈｅ　ｂｕｓ　ｌｅａｖｅｓ」であるときは、距離は10^{-5}となり、部分文が「ｌｅａｖｅ　ｙａｃｈｔ」であるときは、距離は０．５となる。このように生成した、部分文の用例と距離とのデータベースは、データベースメモリ８に記憶される。

【００３４】図６は、以上のように構成された英語処理の音声認識装置の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。図６（ａ）に示すように、「Ｔｈｅｂｕｓ　ｌｅａｖｅｓ　Ｋｙｏｔｏ　ａｔ　１１　ａ．ｍ．」という入力文の音声が入力されたとき、認識結果文として、図６（ｂ）に示すように、「Ｔｈｅｂｕｓ　ｌｅａｖｅｓ　ｙａｃｈｔ　ａｔ　１１　ａ．ｍ．」が得られたとき、すなわち、「Ｋｙｏｔｏ」という地名の固有名詞が「ｙａｃｈｔ」という名詞に誤って認識された場合である。認識結果文における単語間のスコアを図６（ｂ）に示している。さらに、認識結果文に基づいて構文解析したときに、図６（ｃ）に示すように、より小さいスコアに基づいて構文解析結果の構文木が得られ、このときのスコアが得られている。図６（ｃ）における場合を、上記数１に当てはめると、不適格文判定関数の関数値$F_{error}(m)$は次式のようになる。

【００３５】

【数３】

$$F_{error}(m) = (5/4)(10^{-5}+0.5+0.5+10^{-5}) = 1.25\times(1.00002) = 1.250025$$

【００３６】当該例において、上記数３で計算された関数値＝１．２５００２５はしきい値Ｆｔｈを超えているので、それに対応する音声認識候補は音声認識候補から除去される。

【００３７】

【実施例】本発明者は、上述の不適格文検出方法を備えた音声認識装置の有効性を評価するために、以下のごとく実験を行った。ここでは、上述の不適格文判定関数F_{error}が、Ｎ－ｇｒａｍ言語モデルを用いた認識実験における誤認識文と正解文とを区別することが可能かどうかを確認した。具体的には、ｂｉ－ｇｒａｍを用いた認識システムによる誤認識結果文と正解文とを対象に不適格文判定関数F_{error}を算出し、誤認識文と正解文との不適格文判定関数の関数値F_{error}の違いを考察した。正解文では、形態素の数ｍが大きい、つまり部分文が長いほど、文構造が複雑になり構造の曖昧性も低くなるので関数値F_{error}が小さくなり、誤認識文との区別がつきやすくなると想像できる。しかしながら、認識処理の効率化を考えると、なるべく早く、つまり形態素の数ｍが小さい段階の音声認識候補の部分文に対して不適格判定を行ない、不適格文を誤認識文として結果候補から除去することが好ましい。信頼性の高い関数値F_{error}を得るための形態素の数ｍを知るために、誤認識または正解文のｍ番目の形態素までの音声認識候補の部分文に対して関数値F_{error}を計算し、形態素の数ｍを変化させた時の関数値F_{error}の変化も合わせて調べた。実験における音声認識及びデータ条件を表６に示す。

【００３８】

【表６】

音声認識及びデータ条件	

11 12

音響モデル	不特定話者ＨＭ－ｎｅｔ，４０１状態，１０混合分布
言語モデル	単語のｂｉ－ｇｒａｍ
音声認識方式	Ｏｎｅ－ｐａｓｓ　ＤＰ，Ｎ－ｂｅｓｔ探索
bi-gram学習データ	３３６３文、２２２９５４単語
評価データ	学習用データに含まれる４４文、４話者

【００３９】音声認識処理は、統計的言語モデルに単語のｂｉ－ｇｒａｍを使用し、ｏｎｅ－ｐａｓｓ　ＤＰアルゴリズム、Ｎ－ｂｅｓｔ探索型の音声認識システムを用いた。正解文として、表６に示した評価データを用い、誤認識文としては、上記評価データを、表６に示した３種類のＮ－ｇｒａｍを用いた認識システムで認識し、その結果の誤認識文９４文を用いた。図４に、正解文に対する関数値Ｆｅｒｒｏｒの平均値と最大値、及び誤認識文に対する関数値Ｆｅｒｒｏｒを、各形態素数ｍ毎に示す。この図４より、次のことがわかる。

（ａ）正解文については形態素数ｍが長くなるほど、関数値Ｆｅｒｒｏｒの平均値、最大値ともに減少する。

（ｂ）誤認識文においても同様に、形態素数が長くなるほど関数値Ｆｅｒｒｏｒは減少する傾向にあるが、その減少の度合いは正解文に比べて少ない。

【００４０】このことは、左から右への（left-to-rightの）音声認識処理系において、処理した形態素がまだ少ない文の始めの部分では、正解文及び誤認識文の関数値Ｆｅｒｒｏｒに差がなく、不適格文の検出は困難であるが、処理した形態素数が多くなるほど、正解文と誤認識文との関数値Ｆｅｒｒｏｒに差が生じるため、上記しきい値Ｆｔｈを適切に設定することで、不適格文の検出が可能であることを示している。但し、このしきい値Ｆｔｈは一定値ではなく、形態素数ｍを変数とする関数値として定義した方がより有効であることがわかる。例えば、図４中の最大値をしきい値Ｆｔｈとした場合には、このしきい値Ｆｔｈ以上の関数値Ｆｅｒｒｏｒを示す文章は、各々の形態素数ｍの処理を行なっている際に、不適格文と判定することができる。このように文の途中結果から誤認識であると判定できた文の割合は、本実験では全誤認識文中４７．９％（＝４５／９４）であった。以上の結果をまとめると、次のようになる。

（ａ）本不適格文の検出に用いた（１）入力語句と用例との意味的距離、（２）形態素数に対する規則数で表された文構造の複雑さの２つのパラメータは、不適格文を判定するのに有効なパラメータであり、提案した不適格文判定関数Ｆｅｒｒｏｒは、不適格文を検出するのに有効であることがわかった。

（ｂ）不適格文検出の性能は形態素数ｍに依存し、ｍが大きくなるほど、検出性能は上がる。

（ｃ）不適格文判定関数Ｆｅｒｒｏｒのしきい値Ｆｔｈは、形態素数ｍに依存して変えた方が、より効率良く不適格文を検出できる。

【００４１】以上説明したように、本発明によれば、用例との意味的距離を使用することで構文の曖昧性を解消しながら構文を決定していく構文決定手法とを用いて、従来の統計的言語モデルを用いた音声認識の誤認識結果文の不適格性を逐次的に検出する方法を発明した。この方法は、認識結果の部分文に含まれる語句と予め学習された用例との意味的距離と、認識結果の部分文の構文の複雑さとを不適格文の判定要因として使用するものである。そして、様々な単語及び品詞のｂｉ－ｇｒａｍを用いた認識システムの結果を対象に、不適格文の検出を行なった結果、誤認識文と正解文との判定のしきい値Ｆｔｈを適切に設定すれば、誤認識文の約半分を不適格な文として検出可能であることがわかった。

【００４２】従って、音声認識部６は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識するので、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４３】以上の実施形態においては、不適格文判定関数として数１を用いているが、本発明はこれに限らず、以下に示す数４又は数５の不適格文判定関数を用いてもよい。

【数４】

$$F_{error}'(m) = (m/R)\left\{\sum_{i=1}^{R} d(r_i)/R\right\}$$

【数５】

$$F_{error}''(m) = (1/M)\left\{\sum_{i=1}^{R} d(r_i)/R\right\}$$

（m）は、数１の不適格文判定関数Ｆｅｒｒｏｒ（m）に比較して、上記音声認識候補の構文を決定するために用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識候補の構文を決定するために用いた用例の数で除算した値である意味的距離の平均値を計算することを特徴としている。また、数５において、Mは、所定m個の形態素を処理した段階で上記音声認識候補の構文を決定するために用いた用例の規則の中で所定の複数m_a個以上の形態素を含む規則数を表し、ここで、mは好ましくは５以上であって、m_aは好ましくは３である。数５の不適格文判定関数Ｆｅｒｒｏｒ''（m）は、数３の不適格文判定関数Ｆｅｒｒｏｒ'（m）に比較して、（m／Ｒ）に代えて上記規則数Mの逆数を用いたことを特徴とする。これら数４又は数５の不適格文判定関数を用いて音声認識することにより、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４５】

【発明の効果】以上詳述したように本発明に係る請求項１記載の音声認識装置によれば、入力される単語列からなる発声音声文の音声の音声信号に基づいて、所定の統計的言語モデルを参照して上記音声に対して音声認識処理を実行する音声認識手段とを備えた音声認識装置において、上記音声認識手段は、音声認識候補に対して、音声認識候補に対する不適格の度合いを表わす所定の不適格文判定関数を用いて不適格文判定関数の関数値を計算し、計算された関数値が所定のしきい値を超えるときに、当該音声認識候補を除去して音声認識処理を実行する。従って、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４６】また、請求項２記載の音声認識装置においては、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４７】さらに、請求項３記載の音声認識装置においては、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値に音声認識処理の対象となる音声認識候補に含まれる形態素の数を乗算しかつ上記音声認識処理で用いた用例の数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４８】またさらに、請求項４記載の音声認識装置においては、請求項１記載の音声認識装置において、上記不適格文判定関数の関数値は、上記音声認識処理で用いた用例に対応する意味的距離の和を計算し、計算された和を上記音声認識処理で用いた用例の数で除算した値である意味的距離の平均値を計算し、計算された意味的距離の平均値を、所定個の形態素を処理した段階で上記音声認識処理で用いた用例中で所定の複数個以上の形態素を含む用例数で除算した値である。従って、簡便に上記不適格文判定関数の関数値を計算することができ、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【００４９】さらに、請求項５又は６記載の音声認識装置においては、請求項１乃至４のうちの１つに記載の音声認識装置において、上記しきい値は、好ましくは、一定値、もしくは、音声認識処理の対象となる部分文に含まれる形態素の数に依存して変化させる。従って、より有効的に、不適格な誤認識結果を除去することができ、局所的にも大局的にも適格な文を出力でき、従来例に比較して高い音声認識率を得ることができる音声認識装置を提供することができる。

【図面の簡単な説明】

【図１】　本発明に係る一実施形態である音声認識装置のブロック図である。

【図２】　図１の音声認識装置における日本語の単語セットと距離との関係を示す図である。

【図３】　図１の音声認識装置の日本語処理の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。

【図４】　図１の音声認識装置のシミュレーション結果であって、入力された形態素の数に対する判定関数値Ｆｅｒｒｏｒを示すグラフである。

【図５】　図１の音声認識装置における英語の単語セットと距離との関係を示す図である。

【図６】　図１の音声認識装置の英語処理の動作を示す動作図であって、入力文と、認識結果文とその構文木とスコアと、構文解析結果文とその構文木とスコアとを示す動作図である。

【符号の説明】

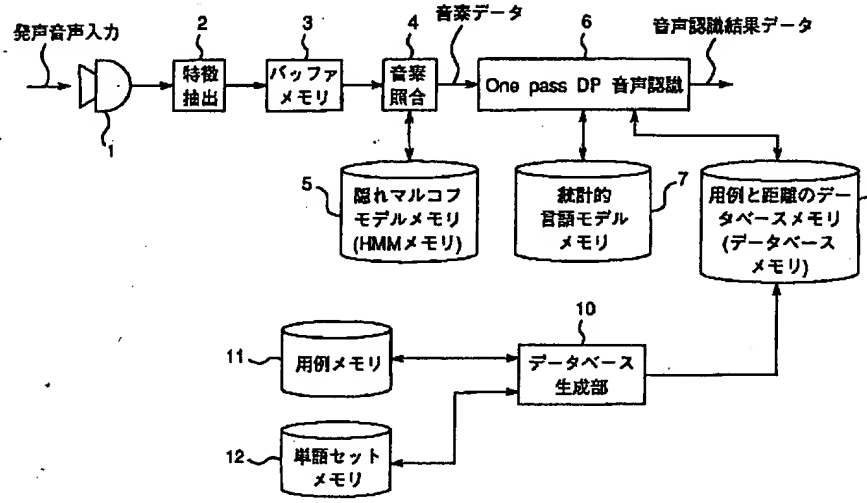
15

1…マイクロホン、
2…特徴抽出部、
3…バッファメモリ、
4…音素照合部、
5…隠れマルコフモデルメモリ（HMMメモリ）、
6…Ｏｎｅ　ｐａｓｓ　ＤＰ音声認識部、

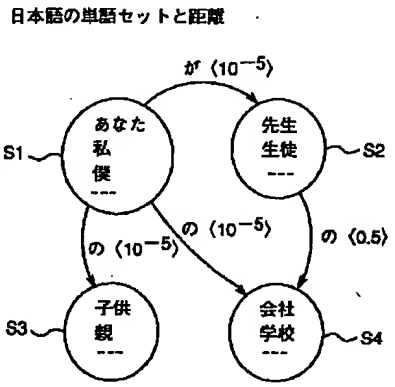
16

7…統計的言語モデルメモリ、
8…用例と距離のデータベースメモリ（データベースメモリ）、
１０…データベース生成部、
１１…用例メモリ、
１２…単語セットメモリ。

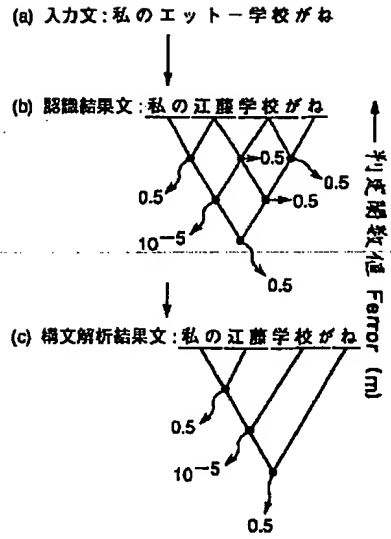
【図１】



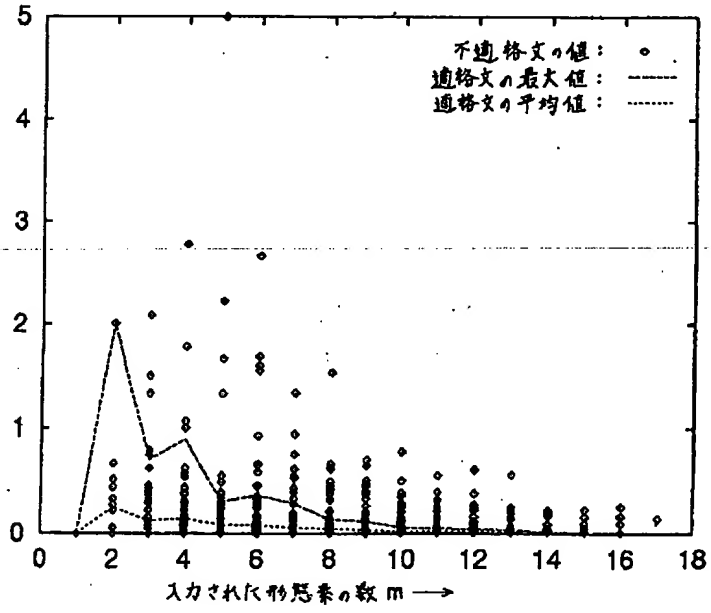
【図２】



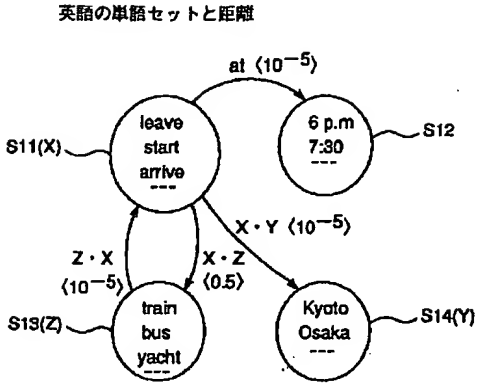
【図３】



【図４】



【図５】

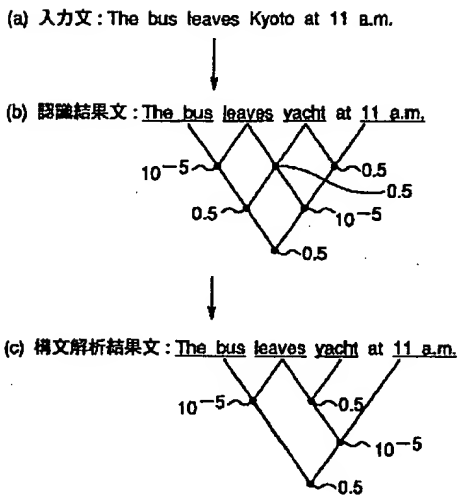

英語の単語セットと距離
at ⟨10⁻⁵⟩
leave
start
arrive
---
6 p.m
7:30

S11(X)
S12
X・Y ⟨10⁻⁵⟩
Z・X
⟨10⁻⁵⟩
X・Z
⟨0.5⟩
train
bus
yacht

Kyoto
Osaka

S13(Z)
S14(Y)

【図６】


(a) 入力文：The bus leaves Kyoto at 11 a.m.
(b) 認識結果文：The bus leaves yacht at 11 a.m.
10⁻⁵
0.5
0.5
0.5
10⁻⁵
0.5
(c) 構文解析結果文：The bus leaves yacht at 11 a.m.
10⁻⁵
0.5
10⁻⁵
0.5